Lockheed Electronics Company, Inc.
A SUBSIDIARY OF LOCKHEED CORPORATION
1830 NASA Road 1, Houston, Texas 77058
Tel. 713-333-5411

Ref: 642-7139
Contract NAS 9-15200
Job Order 73-705-39
143

TECHNICAL MEMORANDUM

ESTIMATION OF PROPORTIONS USING LINEAR MAPS

By

H. F. Walker

Approved By: B L Carroll
for T. C. Minter, Supervisor
Techniques Development Section

August 1978

LEC-12772

ESTIMATION OF PROPORTIONS USING LINEAR MAPS

1. INTRODUCTION

In this memorandum, a statistical population ω is considered which consists of a mixture of members of two statistical populations ω_1 and ω_2 in proportions α_1 and α_2. The problem of estimating α_1 and α_2 on the basis of an unlabeled independent sample of observations on ω in n-dimensional Euclidean space R^n is addressed. It is assumed that μ_1 and μ_2, the respective expected values of observations on ω_1 and ω_2 in R^n, either are known or have been estimated with satisfactory accuracy from a labeled sample of observations on ω_1 and ω_2.

In the following sections it is first shown that, by using a certain derivation, one can obtain unbiased, consistent estimates[1] of α_1 and α_2 from almost any linear map from R^n to R^1. Then that linear map which yields estimates of α_1 and α_2 having minimum variance among all estimates so obtained is sought. A simple expression for the minimum-variance estimates is obtained by following a line of reasoning analogous to that employed in the derivation of the Fisher linear discriminant (ref. 1). The exact evaluation of these estimates requires the use of the (usually unknown) covariance matrix of observations on ω in R^n. In practice, a satisfactory approximation of these estimates can be obtained by using the sample estimate of this covariance matrix instead.

2. ESTIMATION OF PROPORTIONS USING AN ALMOST ARBITRARY LINEAR MAP

Suppose that F is any linear map from R^n to R^1 such that $F(\mu_1) \neq F(\mu_2)$, and suppose that $\chi = \{x_k\}_{k=1,\ldots,N}$ is a sample of independent observations on ω in R^n. It is shown in the following that, from F, one can obtain unbiased, consistent estimates of α_1 and α_2 based on χ. For convenience, write (uniquely) $F(x) - b^T x$ for appropriate $b \epsilon R^n$, and denote by μ the expected value of observations on ω in R^n.

---

[1] For convenience, estimates and their associated estimators are identified.

From the facts that $\mu = \alpha_1\mu_1 + \alpha_2\mu_2$ and $\alpha_1 + \alpha_2 = 1$, one sees that

$$F(\mu) = F(\alpha_1\mu_1 + \alpha_2\mu_2) = F\big(\alpha_1(\mu_1 - \mu_2) + \mu_2\big)$$

$$= \alpha_1[F(\mu_1) - F(\mu_2)] + F(\mu_2)$$

Since $F(\mu_1) \neq F(\mu_2)$, it follows that

$$\alpha_1 = \frac{F(\mu) - F(\mu_2)}{F(\mu_1) - F(\mu_2)}$$

This suggests the estimates

$$\hat{\alpha}_1 = \frac{F(m) - F(\mu_2)}{F(\mu_1) - F(\mu_2)} \tag{1}$$

$$\hat{\alpha}_2 = 1 - \hat{\alpha}_1 \tag{2}$$

where

$$m = \frac{1}{N}\sum_{k=1}^{N} x_k$$

Since m is an unbiased and consistent estimate of μ, it is easily verified that the estimates given by eqs. (1) and (2) are unbiased and consistent. However, if approximate values are given for μ_1 and μ_2, these estimates will be biased accordingly.

3. THE MINIMUM-VARIANCE ESTIMATES

One now determines which linear map F is "best" for use in the estimate given in eq. (1). That linear map for which the estimate given in eq. (1) has minimum variance among all such estimates is considered "best."

The variance of the estimate of eq. (1) is given by

$$\mathrm{Var}(\hat{\alpha}_1) = E\left(|\hat{\alpha}_1 - \alpha_1|^2\right) = E\left(\left|\frac{b^T(m - \mu_2)}{b^T(\mu_1 - \mu_2)} - \alpha_1\right|^2\right)$$

$$= E\left(\frac{1}{|b^T(\mu_1 - \mu_2)|^2} \left| b^Tm - b^T\mu_2 - \alpha_1 b^T\mu_1 + \alpha_1 b^T\mu_2 \right|^2\right)$$

$$= \frac{1}{|b^T(\mu_1 - \mu_2)|^2} E\left(\left|b^Tm - b^T[\alpha_1\mu_1 + (1 - \alpha_1)\mu_2]\right|^2\right)$$

$$= \frac{1}{|b^T(\mu_1 - \mu_2)|^2} E\left(\left|b^T(m - \mu)\right|^2\right)$$

$$= \frac{1}{|b^T(\mu_1 - \mu_2)|^2} b^T E\left((m - \mu)(m - \mu)^T\right)b$$

Now

$$E[(m - \mu)(m - \mu)^T] = E\left(\left[\frac{1}{N}\sum_{k=1}^{N}(X_k - \mu)\right]\left[\frac{1}{N}\sum_{k=1}^{N}(X_k - \mu)\right]^T\right)$$

$$= \frac{1}{N^2}\sum_{k,\ell} E\left((X_k - \mu)(X_\ell - \mu)^T\right)$$

$$= \frac{1}{N^2}\sum_{k=1}^{N} E\left((X_k - \mu)(X_k - \mu)^T\right)$$

$$= \frac{1}{N}\Sigma$$

where $\Sigma = E\left((X - \mu)(X - \mu)^T\right)$ is the covariance matrix of observations on ω in R^n. It follows that

$$\mathrm{Var}(\hat{\alpha}_1) = \frac{1}{N}\frac{b^T\Sigma b}{|b^T(\mu_1 - \mu_2)|^2} \qquad (3)$$

It is evident from eq. (3) that choosing b to minimize $\mathrm{Var}(\hat{\alpha}_1)$ is equivalent to choosing b to maximize the expression

$$\frac{\left|b^T(\mu_1 - \mu_2)\right|^2}{b^T\Sigma b} = \frac{b^T\Sigma\left[\Sigma^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T\right]b}{b^T\Sigma b}$$

Since the operator $\Sigma^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ is symmetric with respect to the inner product $\langle u, v\rangle = u^T\Sigma v$ on R^n, this is a (generalized) Rayleigh quotient. It is maximized when b is an eigenvector of $\Sigma^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ corresponding to the eigenvalue of largest absolute value. Now $\Sigma^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ is an operator of rank 1, and the only eigenvector having an associated eigenvalue which is nonzero is $\Sigma^{-1}(\mu_1 - \mu_2)$. (The eigenvalue associated with this eigenvector is $(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 - \mu_2)$.) It follows that the variance of the estimate in eq. (1) is minimized when F is given by $F(x) = b^Tx$, where $b = \Sigma^{-1}(\mu_1 - \mu_2)$.

With F chosen to minimize the variance of the estimate given by eq. (1), eqs. (1) and (2) can be written as

$$\hat{\alpha}_1 = \frac{(\mu_1 - \mu_2)^T\Sigma^{-1}(m - \mu_2)}{(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 - \mu_2)} \tag{4}$$

and

$$\hat{\alpha}_2 = 1 - \hat{\alpha}_1 \tag{5}$$

The covariance matrix Σ is likely to be unknown in most applications; furthermore, since α_1 and α_2 are unknown, it cannot be determined from a knowledge of μ_1, μ_2, and the covariance matrices for observations in R^n on ω_1 and ω_2. However, when Σ is unknown, it can be replaced in eq. (4) by the sample estimate

$$S = \frac{1}{N - 1}\sum_{k=1}^{N}(x_k - m)(x_k - m)^T$$

to yield the estimates

$$\hat{\alpha}_1 = \frac{(\mu_1 - \mu_2)^T S^{-1}(m - \mu_2)}{(\mu_1 - \mu_2)^T S^{-1}(\mu_1 - \mu_2)} \tag{6}$$

and

$$\hat{\alpha}_2 = 1 - \hat{\alpha}_1 \tag{7}$$

REFERENCE

1. Duda, R. O.; and Hart, P. E.: Pattern Classification and Scene Analysis. John Wiley and Sons, Inc. (New York), 1973.